

SAMPLING PROPERTIES OF THE SPECTRUM AND COHERENCY OF SEQUENCES OF ACTION POTENTIALS

M. R. Jarvis⁽¹⁾ and P. P. Mitra⁽²⁾

February 2, 2008

⁽¹⁾ *Division of Biology, California Institute of Technology, Pasadena, 91125 USA*

⁽²⁾ *Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey, 07974 USA*

Abstract

The spectrum and coherency are useful quantities for characterizing the temporal correlations and functional relations within and between point processes. This paper begins with a review of these quantities, their interpretation and how they may be estimated. A discussion of how to assess the statistical significance of features in these measures is included. In addition, new work is presented which builds on the framework established in the review section. This work investigates how the estimates and their error bars are modified by finite sample sizes. Finite sample corrections are derived based on a doubly stochastic inhomogeneous Poisson process model in which the rate functions are drawn from a low variance Gaussian process. It is found that, in contrast to continuous processes, the variance of the estimators cannot be reduced by smoothing beyond a scale which is set by the number of point events in the interval. Alternatively, the degrees of freedom of the estimators can be thought of as bounded from above by the expected number of point events in the interval. Further new work describing and illustrating a method for detecting the presence of a line in a point process spectrum is also presented, corresponding to the detection of a periodic modulation of the underlying rate. This work demonstrates that a known statistical test, applicable to continuous processes, applies, with little modification, to point process spectra, and is of utility in studying a point process driven by a continuous stimulus. While the material discussed is of general applicability to point processes attention will be confined to sequences of neuronal action potentials (spike trains) which were the motivation for this work.

Keywords and phrases: Spectrum, coherency, coherence, multitaper, lag window, spike train analysis, point processes, finite size effects in spectra, doubly stochastic Poisson process

1 Introduction

The study of spike trains is of central importance to electrophysiology. Often changes in the mean firing rate are studied but there is increasing interest in characterising the temporal structure of spike trains, and the relationships between spike trains, more completely (Gray et al., 1989; Gerstein et al., 1985; Abeles et al., 1983). A natural extension to estimating the rate of neuronal firing is to estimate the autocorrelation and the cross-correlation functions¹. This paper will discuss the frequency domain counterparts of these quantities. Auto- and cross-correlations correspond to spectra and cross spectra respectively. The coherency, which is the normalised cross spectrum, does not in general have a simple time domain counterpart.

The frequency domain has several advantages over the time domain. Firstly often subtle structure can be detected with the frequency domain estimators which is difficult to observe with the time domain estimators. Secondly, the time domain quantities have problems which are associated with sensitivity of the estimators to weak non-stationarity and the non-local nature of the error bars (Brody, 1998). These problems are greatly reduced in the frequency domain. Thirdly, reasonably accurate confidence intervals may be placed on estimates of the second order properties in the frequency domain which permits the statistical significance of features to be assessed. Fourthly, the coherency provides a normalised measure of correlations between time series, in contrast with time domain cross-correlations which are not normalisable by any simple means.

This paper begins by reviewing the population spectrum and coherency for point processes and motivating their use by describing some example applications. Next direct, lag window and multitaper estimators of the spectrum and coherency are presented. The concept of degrees of freedom is introduced and used to obtain large sample error bars for the estimators. Many elements of the work discussed in the review section of this paper can be found in the references (Percival and Walden, 1993; Cox and Lewis, 1966; Brillinger, 1978; Bartlett, 1966). Most of the material in these references is targeted at either spectral analysis of continuous processes or at the analysis of point processes but with less emphasis on spectral analysis. Building on this framework corrections, based on a specific model, will be given for finite sample sizes. These corrections are cast in terms of a reduction in the degrees of freedom of the estimators. For a homogeneous Poisson process the modified degrees of freedom is the harmonic sum of the the asymptotic de-

¹Definitions of these quantities will be given in section 2.3.

degrees of freedom and twice the number of spikes used to construct the estimate. Modifications to this basic result are given for structured spectra and tapered data. A section is included on the treatment of point process spectra which contain lines. A statistical test for the presence of a line in a background of coloured noise is given, and the method for removal of such a line described. An example application to periodic stimulation is given.

2 Population measures and their interpretation

2.1 Counting representation of a spike train

A spike train may be regarded as a point process. If the spike shapes are neglected, it is completely specified by a series of spike times $\{t_i\}$ and the start and end points of the recording interval $[0, T]$. It is convenient to introduce some notation which enables the subsequent formulae to be written in a compact form (Brillinger, 1978). The counting process $N(t)$ is defined as the number of spikes which occur between the start of the interval ($t = 0$) and time t . The counting process has the property that the area beneath it grows as t becomes larger. This is undesirable because it leads to an interval dependent peak at low frequencies in the spectrum. To avoid this a process $\bar{N}(t) = N(t) - \lambda t$, where λ is the mean rate, which has zero mean may be constructed. Note that $d\bar{N}(t) = \bar{N}(t + dt) - \bar{N}(t)$ which is either $1 - \lambda dt$ or $-\lambda dt$ depending on whether or not there is a spike in the interval dt . Thus $d\bar{N}(t)/dt$ is a series of delta functions² with the mean rate subtracted. Figure 1 illustrates the relationship between $N(t)$, $\bar{N}(t)$, and $d\bar{N}(t)/dt$.

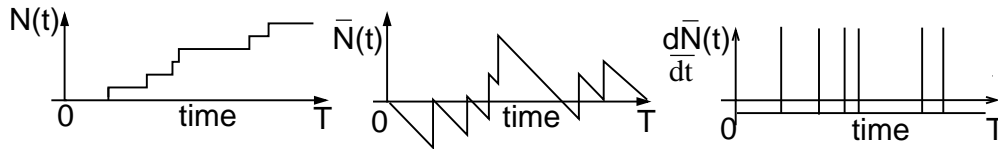


Figure 1: Example illustrating how the processes N , \bar{N} and $d\bar{N}/dt$ relate to each other. The vertical lines in the process $d\bar{N}/dt$ depict delta functions.

²A delta function is a generalized function. It has an area of one beneath it but has zero width and therefore infinite height.

2.2 Stationarity

It will be assumed in what follows that the spike trains are second order stationary. This means that their first and second moments do not depend on the absolute time. In many electrophysiology experiments this is not the case. In awake behaving studies the animal is often trained to perform a highly structured task. Nevertheless it may still be the case that over an appropriately chosen short time window, the statistical properties are changing slowly enough for reasonable estimates of the spectrum and coherency to be obtained. As an example, neurons in primate parietal area PRR exhibit what is known as memory activity during a delayed reach task (Snyder et al., 1997). The mean firing rate of these neurons varies considerably during the task but during the memory period is roughly constant. The assumption of stationarity during the memory period is equivalent to the intuitive notion that there is nothing special about 0.75s into the memory period as compared to say 0.5s. Second order stationarity implies that the mean firing rate (λ) is independent of time and additionally that the autocovariance depends only on the lag (τ) and not on the absolute time.

2.3 Definitions

Equations 1 - 4 give the first and second order moments of a single spike train for a stationary process. The spectrum $S(f)$ is the Fourier transform of the autocovariance function ($\mu(\tau) + \lambda\delta(\tau)$).

$$\frac{E\{dN(t)\}}{dt} = \lambda \quad (1)$$

$$\frac{E\{d\bar{N}(t)\}}{dt} = 0 \quad (2)$$

$$\mu(\tau) + \lambda\delta(\tau) = \frac{E[d\bar{N}(t)d\bar{N}(t+\tau)]}{dtd\tau} \quad (3)$$

$$S(f) = \lambda + \int_{-\infty}^{\infty} \mu(\tau) \exp(-2\pi if\tau) d\tau \quad (4)$$

Where E denotes the expectation operator.

The autocovariance measures how likely it is that a spike will occur at time $t+\tau$ given that one has occurred at time t . Usually $\mu(\tau)$ is estimated rather than the full autocovariance which includes a delta function at zero lag³. However, in order to take the Fourier transform the full autocovariance is required. The inclusion of this delta function leads to a constant offset of the spectrum. This offset is an

³When estimating the autocovariance using a histogram method one generally omits the spike at the start of the interval which would always fall in the bin nearest zero.

important difference between continuous time processes and point processes. The population coherency $\gamma(f)$ is defined in equations 5 - 7.

$$\mu_{ab}(\tau) = \frac{E[d\bar{N}_a(t)d\bar{N}_b(t+\tau)]}{dt d\tau} \quad (5)$$

$$S_{ab}(f) = \int_{-\infty}^{\infty} \mu_{ab}(\tau) \exp(-2\pi i f \tau) d\tau + \lambda_a \delta_{ab} \quad (6)$$

$$\gamma(f) = \frac{S_{12}(f)}{\sqrt{S_{11}(f)S_{22}(f)}} \quad (7)$$

Where indices 1 and 2 denote simultaneously recorded spike trains from different cells.

Unlike the spectrum, which is strictly real and positive, the coherency is a complex quantity. The modulus of the coherency, which is known as the coherence⁴, can only vary between zero and one. This makes coherence particularly attractive for detecting relationships between spike trains as it is insensitive to the mean spike rates.

3 Examples and their interpretation

Before discussing the details regarding how to estimate the spectrum and coherency it will be helpful to motivate them further by considering some simple examples.

3.1 Example population spectra

For a homogeneous Poisson process of constant rate λ the autocovariance is simply $\lambda\delta(\tau)$ and hence the spectrum is a constant equal to the rate λ . At the opposite extreme consider the case where the spikes are spaced by intervals $\Delta\tau$. This is not a stationary process but if a small amount of drift is permitted, so that over an extended period there is nothing special about a given time, it becomes stationary. The spectrum of this process contains sharp lines at integer multiples of $f = \frac{1}{\Delta\tau}$. Due to the drift the higher harmonics will become increasingly blurred and in the high frequency limit the spectrum will tend towards a constant value of the mean rate λ . As a final example consider the case where $\mu(\tau)$ is a negative Gaussian centered on zero τ . This form of $\mu(\tau)$ is consistent with the probability of firing being suppressed after firing⁵. The spectrum of this process will be below λ at

⁴Some authors define coherence as the modulus squared of the coherency.

⁵This need not necessarily correspond to the biophysical refractive period but, it could arise, rather from a characteristic integration time.

low frequencies and will go to a constant value λ at high frequencies. Figure 2 illustrates these different population spectra.

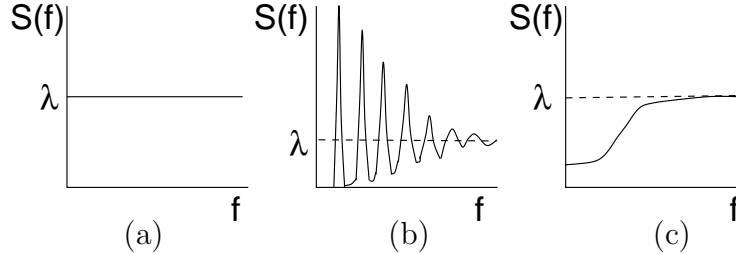


Figure 2: Example population spectra for different types of underlying process. (a) Homogeneous Poisson process with rate λ . (b) Regularly spaced spikes with jitter. (c) Spike trains in which the probability of firing is suppressed immediately after firing.

3.2 Example population coherency

The population coherency of two homogeneous Poisson processes is zero. In contrast if two spike trains are equal then the coherence is one and the phase of the coherency is zero at all frequencies. If two spike trains are identical but offset by a lag $\Delta\tau$ then the coherence will again be one but the phase of the coherency will vary linearly with frequency with a slope proportional to $\Delta\tau$ and given by $\phi(f) = 2\pi f \Delta\tau$.

4 Estimating the spectrum

Section 3 demonstrated that the population spectrum may provide insights into the nature of a spike train. In this section the question of how to estimate the spectrum from a finite section of data will be introduced. In what follows the population quantity λ in the definition of $\overline{N}(t)$ is replaced by a sample estimate $N(T)/T$.

4.1 Direct Spectral Estimators

4.1.1 Definition

A popular, though seriously flawed, method for estimating the spectrum is to take the modulus squared of the Fourier transform of the data $d\overline{N}(t)$. This estimate is known as the Periodogram and is the simplest example of a direct spectral estimator. More generally, a direct spectral estimator is the modulus squared of the Fourier transform of the data but with the data being multiplied by an envelope function $h(t)$, known as a taper (Percival and Walden, 1993). Equations

8 - 10 define the direct estimator. On substituting $\overline{N}(t)$ into equation 9 a form amenable to implementation on a computer is obtained (equation 11). In this form the Fourier transform may be computed rapidly and without the need for the binning of data. Note that equation 10 results in $h(t)$ scaling as $1/\sqrt{T}$ as the sample length is altered. This ensures proper normalization of the Fourier transformation as sample size varies.

$$I^D(f) = |J^D(f)|^2 \quad (8)$$

$$J^D(f) = \int_0^T h(t) e^{-2\pi i f t} d\overline{N}(t) \quad (9)$$

Where,

$$\int_0^T h(t)^2 dt = 1 \quad (10)$$

$$J^D(f) = \sum_{j=1}^{N(T)} h(t_j) e^{-2\pi i f t_j} - \frac{N(T)H(f)}{T} \quad (11)$$

and $H(f)$ is the Fourier transform of the taper.

The direct estimator suffers from bias and variance problems, described below, and is of no practical relevance for a single spike train sample.

4.1.2 Bias

It may not be immediately apparent why the above procedure is an estimate of the spectrum, especially when one is permitted to multiply the data by an arbitrary, albeit normalized, taper. The relation between $I^D(f)$ and the spectrum may be obtained by taking the expectation of equation 8.

$$E\{I^D(f)\} = E\left\{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t)h(t') e^{-2\pi i f(t-t')} d\overline{N}(t)d\overline{N}(t')\right\} \quad (12)$$

Assuming that the integration and expectation operations may be interchanged and substituting equation 3 yields ⁶,

$$E\{I^D(f)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t)h(t') e^{-2\pi i f(t-t')} \{\mu(t-t') + \lambda\delta(t-t')\} dt dt' \quad (13)$$

Which may be rewritten in the Fourier domain as,

$$E\{I^D(f)\} = \int_{-\infty}^{\infty} S(f') |H(f-f')|^2 df' \quad (14)$$

⁶For the moment, we assume that the population quantity λ is known. This is of course not the case in practice, and one employs the estimate $N(T)/T$ as stated before. The effect of this extra uncertainty is given in equation 15.

The expected value of the direct estimator is a convolution of the true spectrum and the modulus squared of the Fourier transform of the taper. The normalization condition on the taper (equation 10) is equivalent to the requirement that the kernel of the convolution has unit area underneath it. Sharp features in the true spectrum will be thus be smeared by an amount which depends on the width of the taper in the frequency domain. If the taper is well localized in the frequency domain the expected value of the direct estimate is close to the true spectrum but if the taper is poorly localized then the expected value of the direct estimator will be incorrect i.e. the direct spectral estimator is biased. There is a fundamental level beyond which the bias cannot be reduced, due to the uncertainty relation forbidding simultaneous localization of a function in the time and frequency domains below a given limit. Since the maximum width of the taper is T the minimum frequency spread is $1/T$ which is known as the Raleigh frequency. Figure 3 shows the smoothing kernel for a rectangular taper and a T of 0.5s. Note that this kernel has large sidelobes which is the primary motivation for using tapering.

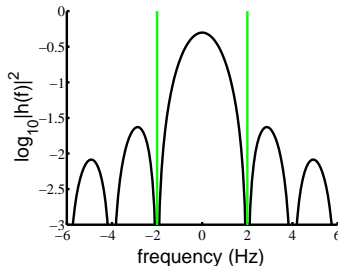


Figure 3: The smoothing kernel $|H(f)|^2$. This is the expected direct estimate of the spectrum in the case of a population spectrum which has a delta function (very sharp feature) at the center frequency. A rectangular taper of length 0.5s was used. Solid vertical lines are drawn at \pm the Raleigh frequency.

In the above argument equation 3 was used in spite of the appearance of the population quantity λ rather than the sample estimate $N(T)/T$ for which equation 12 was defined. A more careful treatment, which includes this correction, leads to an additional term at finite sample sizes in the expectation of the direct spectral estimator at low frequencies. The full expression is given below,

$$E\{I^D(f)\} = \int_{-\infty}^{\infty} S(f')|H(f-f')|^2 df' - |H(f)|^2 S(0)/T \quad (15)$$

In the case of the periodogram, where $h(t) = 1/\sqrt{T}$, the effect is clear since in this case $J^D(0) = 0$ and hence $I^D(0) = 0$ for any set of spike times and any T .

4.1.3 Asymptotic variance

In the previous section it was shown that provided the taper is sufficiently local in frequency the expected value of the direct spectral estimator will be close to the

true spectrum. However, the fact that the estimate is on average close to the true spectrum belies a serious problem with direct spectral estimators, namely that the estimates have very large fluctuations about this mean. The underlying source of this problem is that one is attempting to estimate the value of a function at an infinite number of points using a finite sample of data. The problem manifests itself in the fact that direct spectral estimators are *inconsistent* estimators of the spectrum⁷. In fact it may be shown that, under fairly general assumptions, the estimates are distributed exponentially (or equivalently as $S(f)\chi_2^2/2$) for asymptotic sample sizes (i.e. $T \rightarrow \infty$) (Brillinger, 1972). Figure 4 illustrates that direct spectral estimators are noisy and untrustworthy, a fact emphasised by the observation that the χ_2^2 distribution has a standard deviation equal to its mean. In the next three subsections methods for reducing the variance of direct spectral estimators using different forms of averaging will be discussed.

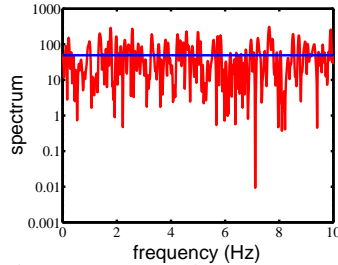


Figure 4: An example of a direct spectral estimate. A 40% cosine taper was used. A sample of duration 20s was drawn from a homogeneous Poisson process with a constant rate of 50 Hz. The population spectrum for this process is flat and is shown by the solid horizontal line. The direct spectral estimate is clearly noisy although on average the correct spectrum is obtained.

4.2 Trial averaging

If there are a number of trials (N_T) available then the variance of the direct estimator may be reduced by trial averaging.

$$I^{DT}(f) = \frac{1}{N_T} \sum_{n=1}^{N_T} I_n^D(f) \quad (16)$$

Where $I_n^D(f)$ is the direct spectral estimate based on the n^{th} trial.

In the large T limit taking the average entails summing N_T independent samples from a χ_2^2 distribution the result of which is distributed as $\chi_{2N_T}^2$. The reduction in variance is inversely proportional to the number of trials corresponding to a reduction in standard deviation which is the familiar factor of $1/\sqrt{N_T}$.

⁷Inconsistent estimators have a finite variance even for an infinite length sample.

At first sight it appears one would be getting something for nothing by breaking a single section of data into N_T segments and treating them as separate trials. This is, of course, not the case. The reason is that if the data is segmented into short length samples, there is loss of frequency resolution proportional to the inverse of segment length. Lag window and multitaper estimators use the information from these independent estimates without artificially segmenting the data.

4.3 Lag Window Estimates

A powerful property of the frequency domain is that, unless two frequencies are very close together, direct estimates of the spectrum of a stationary process at different frequencies are nearly uncorrelated. This property arises when the covariance between frequencies falls off rapidly. If the true spectrum varies slowly over the width of the covariance then the large sample covariance of a direct spectral estimator is given by equation 17.

$$\text{cov}\{I^D(f_1), I^D(f_2)\} \simeq E\{I^D(\bar{f})\}^2 \left| \int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i \Delta f t} dt \right|^2 \quad (17)$$

Where $\bar{f} = (f_1 + f_2)/2$ and $\Delta f = f_1 - f_2$

For $\Delta f = 0$, this expression reduces to the previously mentioned result that the variance of the estimator is equal to the square of the mean. For $\Delta f \gg 1/T$, $\left| \int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i \Delta f t} dt \right|^2 \rightarrow 0$, since $h(t)^2$ is a smooth function with extent T . This implies that $\text{cov}\{I^D(f_1), I^D(f_2)\} \approx 0$ for $|f_1 - f_2| \gg 1/T$. The approximate independence of nearby points means that, if the true spectrum varies slowly enough, then closely spaced points will provide several independent estimates of the same underlying spectrum. This is the motivation for the lag window estimator which is simply a smoothed version of the direct spectral estimator (Percival and Walden, 1993). The lag window estimator is defined in equations 18 and 19.

$$I^{LW}(f) = \int_{-\infty}^{\infty} K(f - f') I^D(f') df' \quad (18)$$

Where,

$$\int_{-\infty}^{\infty} K(f) df = 1 \quad (19)$$

Averaging over trials may be included by using the trial averaged direct spectral estimate I^{DT} (see equation 16) in place of I^D in the above expression. It is assumed that $K(f)$ is a smoothing kernel with reasonable properties.

4.3.1 Bias

The additional smoothing of the lag window kernel modifies the bias properties of the estimator from those expressed in equation 15. The expected value of the

lag window estimator is given by,

$$E\{I^{LW}(f)\} = \int_{-\infty}^{\infty} K(f-f')|H(f'-f'')|^2 S(f'')df'df'' - \frac{S(0)}{T} \int_{-\infty}^{\infty} K(f-f')|H(f')|^2 df' \quad (20)$$

4.3.2 Asymptotic Variance

The large sample variance of this estimator is readily obtained using equation 17.

$$var\{I^{LW}(f)\} = \frac{\xi}{N_T} E\{I^{LW}(f)\}^2 \quad (21)$$

Where,

$$\xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(f)K(f')|\mathcal{H}(f-f')|^2 dfdf' \quad (22)$$

and,

$$\mathcal{H}(f) = \int_{-\infty}^{\infty} h(t)^2 e^{-2\pi i f t} dt \quad (23)$$

Equation 21 includes the reduction in variance due to trial averaging. $1/\xi$ can be interpreted as the effective number of independent estimates beneath the smoothing kernel, as demonstrated by the following qualitative argument. If Δf is the frequency width of the smoothing kernel $K(f)$ and δf is the frequency width of the taper $\mathcal{H}(f)$ then since $K(f) \sim 1/\Delta f$ it follows that $\xi \sim 1/(\Delta f)^2 \int_{\Delta f} \int_{\Delta f} |\mathcal{H}(f-f')|^2 dfdf'$ and hence that $\xi \sim \delta f/\Delta f$.

4.4 Multitaper Estimates

While the lag window estimator is based on the idea that nearby frequencies provide independent estimates, the estimation is not very systematic, since, one should be able to explicitly decorrelate nearby frequencies from the knowledge of the correlations introduced by a finite window size. This is achieved in multitaper spectral estimation. The basic idea of multitaper spectral estimation is to average the spectral estimates from several orthogonal tapers. The orthogonality of the tapers ensures that the estimates are uncorrelated for large samples (consider substituting $h_1(t)h_2(t)$ for $h(t)^2$ in equation 17). A critical question is the choice of a set of orthogonal tapers. A natural choice are the discrete prolate spheroidal sequences (dpss) or Slepian sequences, which are defined by the property that they are maximally localised in frequency. The dpss tapers maximize the spectral concentration defined as;

$$\lambda = \frac{\int_{-W}^W |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df} \quad (24)$$

Where in the time domain $h(t)$ is strictly confined to the interval $[0, T]$.

For given values of W and T there are a finite number of tapers which have concentrations (λ) close to one, and therefore have well controlled bias. This number is known as the Shannon number and is $2WT$. This sets an upper limit on the number of independent estimates that can be obtained for a given amount of spectral smoothing.

A direct multitaper estimate of the spectrum is defined in equation 25.

$$I^{MT}(f) = \frac{1}{K} \sum_{k=0}^{K-1} I_k^D(f) \quad (25)$$

The eigenspectra I_k^D are direct spectral estimates based on tapering the data with the k^{th} dpss function. As previously trial averaging can be included by using I^{DT} rather than I^D . More sophisticated estimates involve adaptive (rather than constant) weighting of the data tapers (Percival and Walden, 1993). Multitaper spectral estimation has been recently shown to be useful for analysing neurobiological time series, both continuous processes (Mitra and Pesaran, 1999) and spike trains (Pesaran et al., 2000).

4.4.1 Bias

The bias for the multitaper estimate is given by equation 15 but with $|H(\cdot)|^2$ replaced by an average over tapers $\frac{1}{K} \sum_{k=0}^{K-1} |H_k(\cdot)|^2$.

4.4.2 Asymptotic Variance

The asymptotic variance of the multitaper estimator, including trial averaging, is given by equation 26.

$$var\{I^{MT}(f)\} = \frac{1}{N_T K} E\{I^{MT}(f)\}^2 \quad (26)$$

4.5 Degrees of freedom

At this point it is useful to introduce the concept of the degrees of freedom (ν_0) of an estimate. The degrees of freedom is twice the number of independent estimates of the spectrum. Degrees of freedom is a useful concept as it permits the expressions for the variance of the different estimators to be written in a common format.

$$var\{I^X(f)\} = \frac{2E\{I^X(f)\}^2}{\nu_0} \quad (27)$$

Where,

X	D	DT	LW	MT
ν_0	2	$2N_T$	$2N_T/\xi$	$2N_TK$

Degrees of freedom is also a useful framework in which to cast both finite size corrections and the confidence limits for the spectra and coherence.

The variance of estimators of the spectrum can be estimated using internal methods such as the bootstrap or jackknife (Efron and Tibshirani, 1993), (Thomson and Chave, 1991). Jackknife estimates can be constructed over trials or over tapers. If ν_0 is large (> 20), then the theoretical and Jackknife variance are in close agreement. If distributional assumptions can be validly made about the point process, theoretical error bars have an important advantage over internal estimates since they enable the understanding of different factors which enter into the variance in order to guide experimental design and data analysis. However Jackknife estimates are less sensitive to failures in distributional assumptions, and this provides them with statistical robustness.

It is conventional to display spectra on a log scale. This is because taking the log of the spectrum stabilizes the variance and leads a distribution which is approximately Gaussian.

4.6 Confidence intervals

The expected values of the estimators and also their variance have been discussed for several different spectral estimators but it is desirable to put confidence intervals on the spectral estimates rather than standard deviations.

As previously mentioned in section 4.2 the averaging of direct spectral estimates from different trials yields, in the large sample limit, estimates which are distributed as $\chi_{2N_T}^2$. In general for the other estimates a well known approximation (Percival and Walden, 1993) is to assume that the estimate is distributed as $\chi_{\nu_0}^2$. Confidence intervals can then be placed on estimates on the basis of this $\chi_{\nu_0}^2$ distribution. The confidence interval applies for the population spectrum $S(f)$ and is obtained from the following argument.

$$P[q_1 \leq \chi_{\nu_0}^2 \leq q_2] = 1 - 2p \quad (28)$$

Where P indicates probability, q_1 is such that $P[\chi_{\nu_0}^2 \leq q_1] = p$ and q_2 is such that $P[\chi_{\nu_0}^2 \geq q_2] = p$. It follows that,

$$P[q_1 \leq \nu_0 I^X(f)/S(f) \leq q_2] = 1 - 2p \quad (29)$$

Hence an approximate $100\% \times (1 - 2p)$ confidence interval for $S(f)$ is given by,

$$P[\nu_0 I^X(f)/q_2 \leq S(f) \leq \nu_0 I^X(f)/q_1] \quad (30)$$

For large ν_0 (> 20) these confidence intervals do not differ substantially from those based on a Gaussian (± 2 standard deviations) but at small ν_0 the difference can be substantial as for these values the $\chi^2_{\nu_0}$ distribution has long tails.

4.7 High Frequency limit

The population spectrum goes to a constant value equal to the rate λ in the high frequency limit. In practice spectra calculated from a finite sample will go to a value close to λ but fluctuations in the number of spikes in the interval will lead to an error in this estimate. For a given sample the spectrum will go the value given by equation 31.

$$I(f \rightarrow \infty) = \frac{1}{N_T K} \sum_{k=0}^{K-1} \sum_{n=1}^{N_T} \sum_{j=1}^{N_n(T)} h_k(t_j^n)^2 \quad (31)$$

Where t_j^n is the j^{th} spike in the n^{th} trial and $N_n(T)$ is the total number of spikes in the n^{th} trial. In the case of direct and lag window estimators the averaging over tapers need not be performed.

This expression yields a value which is typically very close to the sample estimate of the mean rate⁸. It is significant departures from this high frequency limit which are of interest when interpreting the spectrum as these indicate enhancement or suppression relative to a homogeneous Poisson process.

4.8 Choice of estimator, taper and lag window

The preceding section discussed the large sample statistical properties of direct, lag window and multitaper estimates of the spectrum. The choice of which estimator to use remains a contentious one (Percival and Walden, 1993). The multitaper method is the most systematic of the estimators but the lag window estimators should perform almost as well for those spike train spectra which have reasonably small dynamic ranges⁹. However, it is possible to construct spike trains with widely different time scales, which can possess a large dynamic range. In addition, the multitaper technique leads to a simple jackknife procedure by leaving out one data taper in turn. A further important property of the multitaper estimator is that it gives more weight to events at the edges of the time interval and thus ameliorates the arbitrary downweighting of the edges of the data introduced by single tapers.

⁸It is exactly the sample estimate of the mean rate for a rectangular taper.

⁹Dynamic range is a measure of the variation in the spectrum as a function of frequency and is defined as $10 \log_{10}(\frac{\max_f S(f)}{\min_f S(f)})$.

If using the lag window estimator there are many choices available for both the taper and the lag window. The choice of taper is generally not critical provided that the taper goes smoothly to zero at the start and end of the interval. A rectangular taper has particularly large sidelobes in the frequency domain which can lead to significant bias. The choice of lag window is also usually not critical and typically a Gaussian kernel will be satisfactory.

5 Estimating the Coherency

Sample coherency, which may be estimated using equation 32, may be evaluated using any of the previously mentioned spectral estimators. The principle difference is that the direct estimator, in terms of which the other estimators are expressed, is given by equations 33 and 34 rather than 8 and 9.

$$C(f) = I_{12}^X / \sqrt{I_{11}^X I_{22}^X} \quad (32)$$

$$I_{12}^D(f) = J_1^D(f) J_2^D(f)^* \quad (33)$$

$$J_a^D(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f t} d\overline{N}_a(t) \quad (34)$$

Where the $\overline{N}_1(t)$ and $\overline{N}_2(t)$ are simultaneously recorded spike trains from two different cells and X denotes the type of spectral estimator. Possible choices of estimator X include; D direct, DT trial averaged direct, LW lag window or MT multitaper.

Lag window and multitaper coherency estimates may be constructed by substituting $I_{12}^D(\cdot)$ in place of $I^D(\cdot)$ in equations 18 and 25. The estimates are biased over a frequency range equal to the width of the smoothing although the exact form for the bias is difficult to evaluate.

5.1 Confidence limits for the Coherence

The treatment of error bars is somewhat different between the spectrum and the coherency, since the coherency is a complex quantity. Usually one is interested in establishing whether there is significant coherence in a given frequency band. In order to do this the sample coherence should be tested against the null hypothesis of zero population coherence. The distribution of the sample coherence under this null hypothesis is given below.

$$P(|C|) = (\nu_0 - 2)|C|(1 - |C|^2)^{(\nu_0/2 - 2)} \quad 0 \leq |C| \leq 1 \quad (35)$$

A derivation of this result is given in (Hannan, 1970). In outline the method is to rewrite the coherence in such a way that it is equivalent to a multiple correlation

coefficient (Anderson, 1984). The distribution of a multiple correlation coefficient is then a known result from multivariate statistics. In the case of coherence estimates based on lag window estimators the appropriate ν_0 may be used although this is only approximately valid because this method of derivation assumes integer $\nu_0/2$.

It is straightforward to calculate a confidence level based on this distribution. The coherence will only exceed $\sqrt{1 - p^{1/(\nu_0/2-1)}}$ in $p \times 100\%$ of experiments. In addition it is notable that the quantity $(\nu_0/2 - 1)|C|^2/(1 - |C|^2)$ is distributed as F_{2,ν_0-2} under this null hypothesis. It is useful to apply a transformation to the coherence before plotting it which aids in the assessment of significance. The variable $q = \sqrt{-(\nu_0 - 2)\log(1 - |C|^2)}$ has a Raleigh distribution which has density $p(q) = qe^{-q^2/2}$. This density function does not depend on ν_0 and furthermore has a tail which closely resembles a Gaussian. For certain values of a fitting parameter¹⁰ β , a further linear transformation $r = \beta(q - \beta)$ leads to a distribution which closely resembles a standard normal Gaussian for $r > 2$. This means that for $r > 2$ one can interpret r as the number of standard deviations by which the coherence exceeds that expected under the null hypothesis.

5.2 Confidence Limits for the Phase of the Coherency

If there is no population coherency then the phase of the sample coherency is distributed uniformly. If, however, there is population coherency then the distribution of the sample phase is approximately Gaussian provided that the tails of the Gaussian do not extend beyond a width 2π . An approximate 95% confidence interval for the phase (Rosenberg et al., 1989; Brillinger, 1974) is given below.

$$\hat{\phi}(f) \pm 2\sqrt{\frac{2}{\nu_0} \left(\frac{1}{|C(f)|^2} - 1 \right)} \quad (36)$$

Where $\hat{\phi}(f)$, the sample estimate of the coherency phase, is evaluated using $\tan^{-1}\{Im(C)/Re(C)\}$.

6 Finite Size Effects

In the preceding sections error bars were given for estimators of the spectrum and the coherence. However these error bars were based on large sample sizes (they apply asymptotically as $T \rightarrow \infty$). Neurophysiological data are not collected in this regime and, particularly in awake behaving studies where data is often sparse, corrections arising at small T are potentially important. In order to estimate the size of these corrections a particular model for the point process is required.

¹⁰A reasonable choice for β is 23/20.

The model studied was chosen primarily for its analytical tractability while still maintaining a non-trivial spectrum.

The model and the final results will be presented here but the details of the analysis are reserved until appendix A. The model is a doubly stochastic inhomogeneous Poisson process with a Gaussian rate function. A specific realization of a spike train is generated from the model in the following manner. Firstly a population spectrum $S_G(f)$ is specified. From this a realization of a zero mean Gaussian process $\lambda_G(f)$ is generated. A constant λ , the mean rate, is then added to this realization. This function is then considered to be the rate function for an inhomogeneous Poisson process. A realization of this inhomogeneous Poisson process is then generated. A schematic of the model is shown in figure 5.

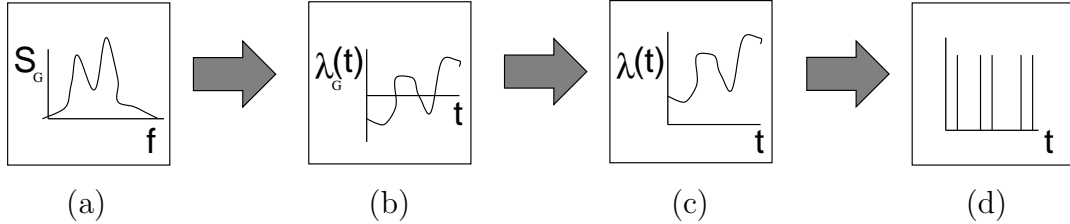


Figure 5: Schematic illustrating the model for which finite size corrections to the asymptotic error bars will be evaluated. (a) A spectrum $S_G(f)$ is defined (b) A realization $\lambda_G(t)$ is drawn from a Gaussian process with this spectrum (c) The mean rate λ is added to $\lambda_G(t)$ to obtain $\lambda(t)$ (d) This rate function is used to generate a realization of an inhomogeneous Poisson process yielding a set of spike times.

Technically this is not a valid process because the rate function $\lambda(t)$ may be negative. However, if the area underneath the spectrum is small enough then the fluctuations about the mean rate seldom cross zero and corrections due to this effect are negligible. In addition large violations of this area constraint have been tested by Monte Carlo simulation and the results still apply to a good approximation.

An important feature of this model is that the population spectrum of the spike trains is simply the spectrum of the inhomogeneous Poisson process rate function plus an offset equal to the mean rate¹¹. The spectrum of the rate function is a positive real quantity and therefore within this model the population spectrum cannot be less than the mean rate at any frequency. Intuitively, the reason for this is that the process must be more variable than a homogeneous Poisson process at all frequencies.

To make the nature of the result clear a simplified version is given in equation

¹¹This result does not depend on the Gaussian assumption.

37. This version is for the particular case of a homogeneous Poisson process (which has a flat population spectrum) and a rectangular taper¹².

$$\text{var}\{I^X(f)\} = \lambda^2 \left[\frac{2}{\nu_0} + \frac{1}{N_T T \lambda} \right] \quad (37)$$

Where λ is the mean rate.

A sample based estimate of $N_T T \lambda$ is the total number of spikes over all trials. It is to be noted that finite size effects reduce the degrees of freedom. This result implies that there is a point beyond which additional smoothing does not decrease the variance further and this point is approximately when ν_0 is equal to twice the total number of spikes. The full result is given in equations 38 - 43.

$$\text{var}\{I^X(f)\} = \frac{2E\{I^X(f)\}^2}{\nu(f)} \quad (38)$$

$$\frac{1}{\nu(f)} = \frac{1}{\nu_0} + \frac{C_h^X \Phi(f)}{2T N_T E\{I^X(f)\}^2} \quad (39)$$

Where,

$$C_h^X = \begin{cases} \int_0^1 f(t)^4 dt & \text{If } X = \text{LW, D or DT} \\ \frac{1}{K^2} \sum_{k,k'} \int_0^1 f_k(t)^2 f_{k'}(t)^2 dt & \text{If } X = \text{MT} \end{cases} \quad (40)$$

$$f(t/T) = \sqrt{T} h(t) \quad (41)$$

and,

$$\Phi(f) = \lambda_{hf} + 4[E\{I^X(f)\} - \lambda_{hf}] + 2[E\{I^X(0)\} - \lambda_{hf}] + [E\{I^X(2f)\} - \lambda_{hf}] \quad (42)$$

$$\lambda_{hf} = E\{I^X(f \rightarrow \infty)\} \quad (43)$$

C_h^X is a constant of order unity which depends on the taper. When a taper is used to control bias some of the spikes are effectively disregarded and this has an effect on the size of the correction. The function $f(t)$ has the same form as the taper $h(t)$ but is defined for the interval $[0, 1]$. C_h^X is the integral of the fourth power of f and obtains its minimum value of one for a rectangular taper. It is typically between 1 and 2 for other tapers. In the multitaper case cross terms between tapers are included.

Equation 42 describes how the finite size correction depends on the structure of the spectrum. $\Phi(f)$ is the sum of four terms. The first term is the only

¹²The expression also holds approximately for the multitaper estimate provided all tapers up to the Shannon limit are used.

term which is present for a flat spectrum. The second term is a correction which depends on the spectrum at the frequency being considered. The next two terms depend on the spectrum at zero frequency and the spectrum at twice the frequency being considered. The latter three terms all depend on the difference between the spectrum at some frequency and the high frequency limit. Equation 42 applies provided that the spike train is well described by the model. However, this is not necessarily the case and a suppression of the spectrum, which cannot be described by the model, often occurs at low frequencies¹³. In the event that there is a significant suppression of the spectrum $\Phi(f)$ may become small or even negative. To avoid this a modified form for $\Phi(f)$ may be used which prevents this.

$$\begin{aligned} \Phi(f) = \lambda_{hf} + 4\max([E\{I^X(f)\} - \lambda_{hf}], 0) &+ 2\max([E\{I^X(0)\} - \lambda_{hf}], 0)... \\ &+ \max([E\{I^X(2f)\} - \lambda_{hf}], 0) \quad (44) \end{aligned}$$

The above modification to the result is somewhat *ad hoc* so Monte Carlo simulations of spike trains with enforced refractory periods have been performed to test its validity. These simulations demonstrated that, although the correction derived using 44 was significantly different from that obtained from the Monte Carlo simulations in the region of the suppression, equation 44 provided a pessimistic estimate in all cases studied. This increases confidence that applying finite size corrections using equation 44 will provide reasonable error bars for small samples.

Equation 39 gives the finite size correction in terms of a reduction in ν_0 . The new $\nu(f)$ may be used to put confidence intervals on the results, as described in section 4.6, although the accuracy of the χ^2_ν assumption will be reduced. In the case of the coherence an indication of the correction to the confidence level can be obtained by using the smaller of the two $\nu(f)$ from the spike train spectra to calculate the confidence level using equation 35. In all cases if the effect being observed only achieves significance by an amount which is of the same order as the finite size correction then it is recommended that more data be collected.

7 Experimental Design

Often it is useful to know in advance how many trials or how long a time interval one needs in order to resolve features of a certain size in the spectrum or the coherence. To do this one needs to estimate the asymptotic degrees of freedom ν_0 . This depends on the size of feature to be resolved α , the significance level for which confidence intervals will be calculated p and the fraction of experiments which will achieve significance \mathcal{P} . In addition the reduction in the degrees of freedom due to finite size effect depends on the total number of spikes N_s and also C_h (see section 6).

¹³Note that any spike train spectra displaying significant suppression below the mean firing rate can immediately rule out the inhomogeneous Poisson process model.

An estimate of ν_0 may be obtained in two stages. Firstly α, p and \mathcal{P} are specified and used to calculate a degrees of freedom ν . Secondly the asymptotic degrees of freedom ν_0 is estimated using ν , N_s and C_h . The feature size $\alpha = (S - \lambda)/\lambda$ is the minimum size of feature which the experimenter is content to resolve. For example, a value of 0.5 indicates that where the population spectrum exceeds 1.5λ the feature will be resolved. The significance level should be set to the same value that will be used for calculating the confidence interval for the spectrum, typically be 0.05. For a given p there is some probability \mathcal{P} that an experiment will achieve significance. To calculate ν one begins with a guess ν_g . Then q_1 is chosen such that $P[\chi_{\nu_g}^2 \geq q_1] = p/2$. On the basis of this one then evaluates $\mathcal{P}_i = 1 - \Phi[q_1/(1 + \alpha)]$ where Φ is the cumulative $\chi_{\nu_g}^2$ distribution¹⁴. If \mathcal{P}_i is equal to the specified fraction \mathcal{P} then $\nu = \nu_g$ otherwise a different ν_g is chosen. This procedure is readily implemented as a minimization of $(\mathcal{P} - \mathcal{P}_i(\nu_g))^2$ on a computer. Having obtained ν one can estimate ν_0 using the following expression.

$$\frac{1}{\nu_0} = \frac{1}{\nu} - \frac{C_h [1 + 4\alpha]}{2N_s [1 + \alpha]^2} \quad (45)$$

Where the 4α is omitted from the numerator if $\alpha < 0$.

Figure 6 illustrates example design curves generated using this method. These curves show the asymptotic degrees of freedom as a function of feature size for different total numbers of spikes.

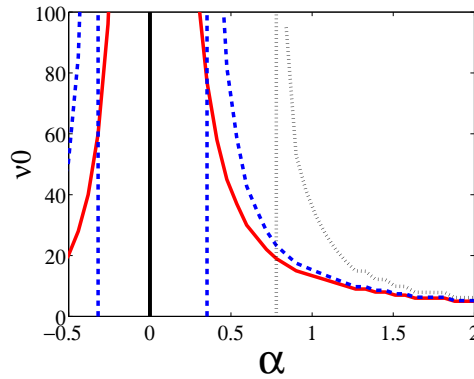


Figure 6: Example design curves for the case when $p = 0.05$, $\mathcal{P} = 0.5$ and $C_h = 1.5$. The three curves correspond to $N_s = \infty$ (solid), $N_s = 100$ (dashed), $N_s = 50$ (dotted).

The existence of a region bounded by vertical asymptotes implies that as long as the total number of measured spikes is finite, modulations in the spectrum

¹⁴These formulae apply for $\alpha > 0$. If $\alpha < 0$ then $P[\chi_{\nu_g}^2 \leq q_1] = p/2$ and $\mathcal{P}_i = \Phi[q_1/(1 + \alpha)]$ should be used.

below a certain level cannot be detected no matter how much the spectrum is smoothed. These curves may be used to design experiments capable of resolving spectral features of a certain size.

In the case of the coherence one calculates how many degrees of freedom are required for the confidence line to lie at a certain level as described in section 5.1.

8 Line Spectra

One of the assumptions underlying the estimation of spectra is that the population spectrum varies slowly over the smoothing width (W for multitaper estimators). While this is often the case there are situations in which the spectrum contains very sharp features which are better approximated by lines than by a continuous spectrum. This corresponds to periodic modulations of the underlying rate, such as when a periodic stimulus train is presented. In such situations it is useful to be able to test for the presence of a line in a background of colored noise (i.e. in a locally smooth but otherwise arbitrary continuous population spectrum). Such a test has been previously developed, in the context of multitaper estimation, for continuous processes (Thomson, 1982) and in the following section the analogous development for point processes is presented.

8.1 F-test for point processes

A line in the spectrum has an exactly defined frequency and consequently the process $N(t)$ has a non-zero first moment. The natural model in the case of a single line is given by equation 46.

$$E\{dN(t)\}/dt = \lambda_0 + \lambda_1 \cos(2\pi f_1 t + \phi) \quad (46)$$

A zero mean process (\overline{N}) may be constructed by subtraction of an estimate of $\lambda_0 t$. Provided that the product of the line frequency (f_1) and the sample duration (T) is much greater than one the sample quantity $N(T)/T$ is an approximately unbiased estimate of λ_0 . The resultant zero mean process \overline{N} has a Fourier transform which has a non-zero expectation.

$$J_k(f) = \int_{-\infty}^{\infty} h_k(t) e^{-2\pi i f t} d\overline{N}(t) \quad (47)$$

$$E\{J_k(f)\} = c_1 H_k(f - f_1) + c_1^* H_k(f + f_1) \quad (48)$$

Where,

$$c_1 = \lambda_1 e^{i\phi}/2 \quad (49)$$

In the case where $f > 0$ and $f_1 > W$,

$$E\{J_k(f)\} \simeq c_1 H_k(f - f_1) \quad (50)$$

The estimates of $J_k(f_1)$ from different tapers provide a set of uncorrelated estimates of $c_1 H_k(0)$. It is hence possible to estimate the value of c_1 by complex regression.

$$\hat{c}_1 = \frac{\sum_k J_k(f_1) H_k(0)}{\sum_k |H_k(0)|^2} \quad (51)$$

Under the null hypothesis that there is no line in the spectrum ($c_1 = 0$) it may readily be shown that $E\{\hat{c}_1\} = 0$ and $var\{\hat{c}_1\} = S(f_1)/\sum_k |H_k(0)|^2$. The residual spectrum¹⁵, which has the line removed, may be estimated using equation 52.

$$\hat{S}(f) = \frac{1}{K} \sum_k |J_k(f) - \hat{c}_1 H_k(f - f_1)|^2 \quad (52)$$

In the large sample limit the distributions of both \hat{c}_1 and $\hat{S}(f_1)$ are known (Percival and Walden, 1993) and may be used to derive relation 53.

$$\frac{|\hat{c}_1|^2 \sum_k |H_k(0)|^2 (K-1)}{\sum_k |J_k(f_1) - \hat{c}_1 H_k(0)|^2} \doteq F_{2,2(K-1)} \quad (53)$$

Where \doteq denotes ‘is distributed as’.

The null hypothesis may be tested using this relation and, if rejected, the line can be removed using equation 52 to estimate the residual spectrum. It is worth noting that although relation 53 was derived for large samples the test is remarkably robust as the sample size is decreased. Numerical tests indicate that the tail of the F distribution is well reproduced even in situations where there are as low as 5 spikes in total.

8.2 Periodic Stimulation

A common paradigm in neurobiology where line spectra are particularly important is that of periodic stimulation. When a neuron is driven by a periodic stimulation of frequency f_1 the spectrum may contain lines at any of the harmonics nf_1 . Provided that $f_1 > 2W$ the analysis of section 8.1 applies with each harmonic being separately tested for significance.

The first moment of the process, which has period $1/f_1$, is given by equation 54 and may be estimated using \hat{c}_n .

$$\lambda(t) = \lambda_0 + \sum_n \lambda_n \cos(2\pi n f_1 t + \phi_n) \quad (54)$$

Where $\lambda_n = 2|c_n|$, $\phi_n = \tan^{-1}\{Im(c_n)/Re(c_n)\}$, the sum is taken over all the significant coefficients.

¹⁵It is also possible to estimate a residual coherency. In order to do this one uses a residual cross-spectrum $\hat{S}_{xy}(f) = \frac{1}{K} \sum_k (J_k^x(f) - \hat{c}_1^x H_k(f - f_1))^* (J_k^y(f) - \hat{c}_1^y H_k(f - f_1))$, together with the residual spectra to evaluate the usual expression for coherency.

This rate function $\lambda(t)$ is the average response to a single stimulus or impulse response. The coefficients c_n are the Fourier series representation of $\lambda(t)$.

8.3 Error Bars

It is possible to put confidence intervals on both the modulus and the phase of the coefficients \hat{c}_n . For large samples (> 10 spikes) the real and imaginary parts of \hat{c}_n are distributed as independent Gaussians each with standard deviation $\sigma_n = \sqrt{S(nf_1)/(2\sum_k |H_k(0)|^2)}$. For $c_n/\sigma_n > 3$ the distribution of $|\hat{c}_n|$ is well approximated by a Gaussian centered on $|c_n|$ and with standard deviation σ_n . In addition the estimated phase angle ($\hat{\phi}_n$) is also almost Gaussian with mean ϕ_n and standard deviation $\sigma_n/|c_n|$. Approximate error bars or confidence intervals may be obtained using a sample based estimate of σ_n , $\hat{\sigma}_n = \sqrt{\hat{S}(nf_1)/(2\sum_k |H_k(0)|^2)}$.

Estimating error bars for the impulse response function is more involved due to their non-local nature (if one of the Fourier coefficients is varied the impulse response function changes everywhere). It is therefore of interest to estimate a global confidence interval, defined as any interval such that the probability of the function crossing the interval *anywhere* is some predefined probability. A method for estimating a global confidence band is detailed in (Sun and Loader, 1994) and outlined here. First a basis vector $\Phi(t)$ is constructed.

$$\Phi(t) = \begin{bmatrix} \hat{\sigma}_1 \cos(2\pi f_1 t) \\ \vdots \\ \hat{\sigma}_N \cos(2\pi f_N t) \\ \hat{\sigma}_1 \sin(2\pi f_1 t) \\ \vdots \\ \hat{\sigma}_N \sin(2\pi f_N t) \end{bmatrix} \quad (55)$$

Where N is the total number of harmonics.

The elements of this vector have unit variance and a standard approximation may be applied.

$$P(\sup |\lambda(t) - E\{\lambda(t)\}| > c \|\Phi(t)\|) \leq 2(1 - N(c)) + (k/\pi)e^{-c^2/2} \quad (56)$$

Where \sup is the maximum value of its operand, $\|\Phi(t)\|$ denotes the length of vector $\Phi(t)$, $N(c)$ is the cumulative standard normal distribution and k is a constant. k may be evaluated by constructing the $2 \times N$ matrix $X(t) = [\Phi(t) d\Phi(t)/dt]$, forming its QR decomposition (Press et al., 1992) and then evaluating $k = \int_0^T |R_{22}(t)/R_{11}(t)| dt$.

Confidence intervals for the residual spectrum are calculated in the usual manner (using χ^2_ν) although at the line frequencies the interval is slightly broadened due to the loss of a degree of freedom incurred by estimation of c_n . Section 11 contains an example application of the methods described in this section.

9 Example Spectra

Figure 7 is a spectrum calculated from data collected from a single cell recorded from area PRR in the parietal cortex of an awake behaving monkey during a delayed memory reach task (Snyder et al., 1997). The spectrum is calculated over an interval of 0.5s during which the firing rate is reasonably stationary and is averaged over 5 trials. The spectrum shows two features which achieve significance. There is enhancement of the spectrum in the frequency band 20-40 Hz indicating the presence of an underlying broad band oscillatory mode in the neuronal firing rate. In addition there is suppression of the spectrum at low frequencies. As discussed previously a suppression of this sort is consistent with an effective refractory period during which the neuron is less likely to fire. Care must be taken at low frequencies since at frequencies comparable to the smoothing width the spectrum is particularly sensitive to any non-stationarity in the data.

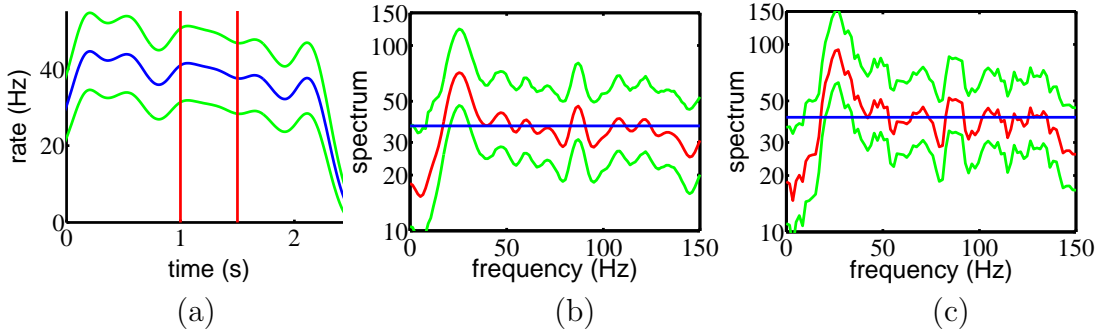


Figure 7: (a) Gaussian kernel (100 ms width) smoothed firing rate with 2σ error bars based on a stationarity assumption. The vertical lines indicate the period over which the spectrum was calculated. A light is flashed at time zero and the spectrum is evaluated over the interval when the monkey is required to remember the target location. (b) The spectrum evaluated over this interval using a lag window estimator with a 40% cosine taper and a Gaussian lag window of width 3.5 Hz. 95% confidence limits are shown with the finite size correction included (this typically resulted in a decrease in $\nu(f)$ from about 50 to 36). The horizontal line indicates the high frequency limit. (c) The same spectrum evaluated using a multitaper estimator. A bandwidth (W) of 5 Hz was used allowing 5 tapers. Both estimators have the same degrees of freedom.

10 Example Coherency

To illustrate the estimation of coherency simulated spike trains were generated from a coupled doubly stochastic Poisson process. For a given trial a pair of rate functions were drawn from a Gaussian process. The realizations share a coherent mode which is linearly mixed into the rates of both cells. These coupled rate functions are then used to independently draw a realization of an inhomogeneous Poisson process for each cell. Using this method 15 trials of duration 0.5s were generated. The coherent mode was set such that the population coherence was a Gaussian of height 0.35 and standard deviation 5 Hz centered on 20 Hz. The phase of this mode was set to 180° . Figure 8 indicates that this coherent mode is reasonably estimated.

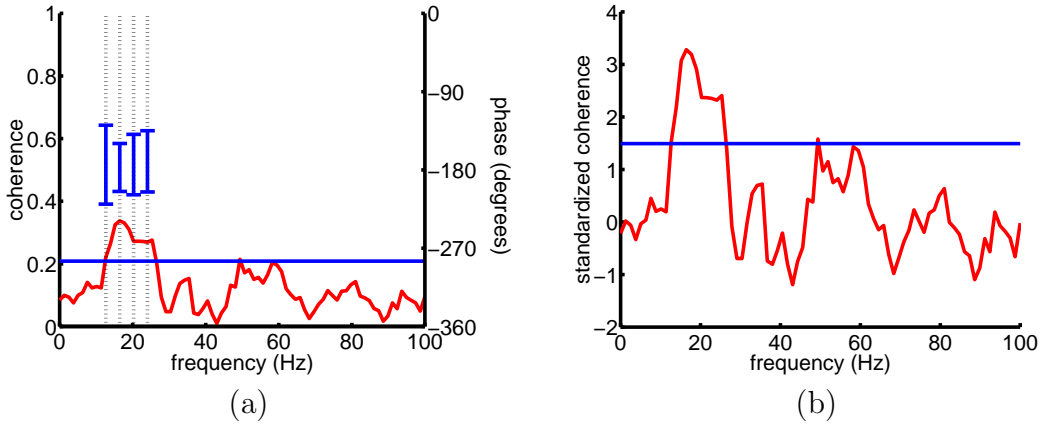


Figure 8: (a) Coherence (left axis) and the phase of the coherency (right axis). Fifteen trials of 0.5s duration were simulated using a doubly stochastic Poisson process as described in the text. A multitaper estimator with a smoothing width of 7 Hz was used. Finite size corrections were used and resulted in 25% reduction in the degrees of freedom. A horizontal line has been drawn at the 95% confidence level under the null hypothesis of no coherency. Where the null hypothesis is rejected the phase of the coherency is estimated and shown with an approximate 95% confidence interval. (b) The standardized coherence is a transformation which maps the null distribution onto an approximately standard normal variate (as described in section 5.1). The estimated coherence at 20 Hz would therefore lie at three standard deviations if there were no population coherence.

11 Example Periodic Stimulation

An example of an analysis of a periodic stimulus paradigm is shown in figure 9. The data is a single cell recording collected from the barrel cortex of an awake

behaving rat during periodic whisker stimulation at 5.5 Hz (Sachdev et al., 1999). There is a single trial of duration 50s.

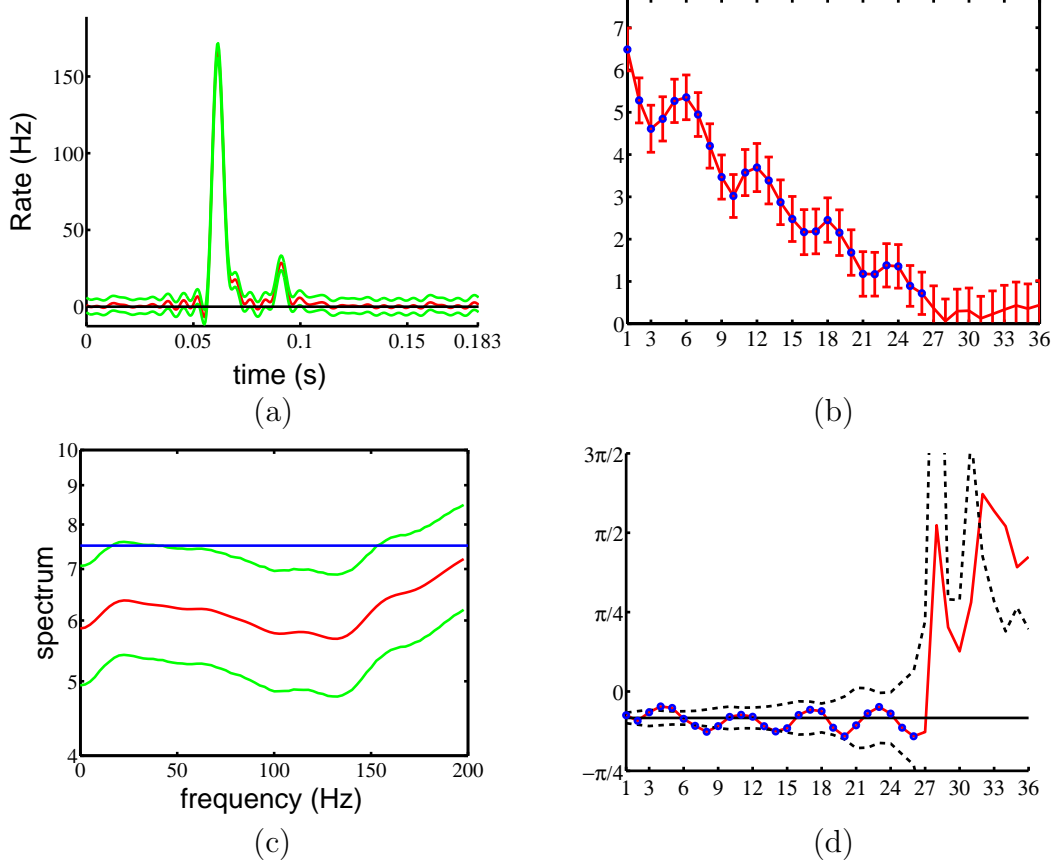


Figure 9: Response to a periodic stimulation of frequency 5.5 Hz. (a) Impulse response function with global 95% confidence interval (b) $|c_n|$ versus index n with 95% confidence interval. Dots indicate points which achieved significance in the F-test. (c) Residual spectrum with finite size corrected confidence interval. A multitaper spectrum with 100 tapers and a bandwidth of 1.5 Hz was used initially to avoid overlap of harmonics. This spectrum was then further smoothed using a Gaussian lag window with standard deviation 9 Hz. (d) The coefficient phases $\hat{\phi}_n$ (in radians) versus index n after subtraction of a fitted straight line of gradient $2\pi/3 \pm 0.01$. The black dashed lines are a 95% confidence interval about zero.

The estimated impulse response function $\hat{\lambda}(t)$ is seen to have two distinct sharp peaks outside of which the response does not differ significantly from zero. The moduli of the Fourier coefficients are significant out to $n = 25$. This automatically sets the smoothing of $\hat{\lambda}(t)$ as structure on a time scale of less than $1/(25 \times 5.5) = 7$ ms does not achieve significance. Note that the coefficients are enhanced at multiples of 6 (i.e. ~ 33 Hz) which comes from having two peaks in the time

$$\boxed{\begin{aligned} J_a^{kn}(f) &= \int_0^T h_k(t) e^{-2\pi i f t} d\bar{N}_a^n(t) \\ I_{ab}^{kn}(f) &= J_a^{kn}(f) J_b^{kn*}(f) \end{aligned}}$$

Table 1: The basic direct spectral estimator in terms of which the other estimators can be written. For clarity the superscript D on the direct spectral estimate has been omitted. The index n labels trials, index k labels tapers, and indices a and b label cells.

domain $\lambda(t)$ which are separated by ~ 30 ms. The phase of the coefficients closely follows a straight line but there is a small periodic deviation from this line which is again at index multiples of 6. The gradient of the straight line depends on the time delay of the response. The residual spectrum was calculated by first evaluating a multitaper estimate from which the significant harmonics were removed. This spectrum had a bandwidth of 1.5 Hz chosen to avoid overlap of the harmonics leading to the multitaper estimate being undersmoothed. A further smoothing was performed using a lag window¹⁶ and the resultant spectrum, displays a slight but significant suppression relative to a Poisson process out to almost 200 Hz. Such a spectrum is characteristic of a short time scale refractive period. The residual spectrum is particularly useful because rate non-stationarity has been removed.

12 Summary

It is the belief of the authors that spectral analysis is a fruitful and under exploited analysis technique for spike trains. In this paper an attempt has been made to collect the machinery necessary for performing spectral analysis on spike train data into a single document. Starting from the population definitions the statistical properties of estimators of the spectrum and coherency have been reviewed. Estimation methods for both continuous spectra and spectra which contain lines have been included. In addition new corrections to asymptotic error bars have been presented which increase confidence in applying spectral techniques in practical situations where data is often sparse. Tables 1 to 5 summarize the important formulae. Matlab software implementing the methods discussed in this paper is available from xxx.lanl.gov/archive/neuro-sys.

¹⁶The previous theory developed for lag window estimators applies to this hybrid estimator with $|H(\cdot)|^2$ replaced by $\frac{1}{K} \sum_{k=0}^{K-1} |H_k(\cdot)|^2$ in equation 20 and $|\mathcal{H}(\cdot)|^2$ replaced by $\frac{1}{K} \sum_{k=0}^{K-1} |\mathcal{H}_k(\cdot)|^2$ in equation 22.

X	$I_{ab}^X(f)$	Eq.	ν_0
D	$I_{ab}^{01}(f)$	[8]	2
DT	$\frac{1}{N_T} \sum_{n=1}^{N_T} I_{ab}^{0n}(f)$	[16]	$2N_T$
LW	$\frac{1}{N_T} \sum_{n=1}^{N_T} \int_{-\infty}^{\infty} K(f-f') I_{ab}^{0n}(f') df'$	[18]	$2N_T/\xi$
MT	$\frac{1}{N_T K} \sum_{n=1}^{N_T} \sum_{k=0}^{K-1} I_{ab}^{kn}(f)$	[25]	$2N_T K$

Table 2: The different estimators and the large sample degrees of freedom ν_0 of estimates of the spectrum ($ab = 11$). The indices on the I_{ab}^{kn} are as follows. ab label the cells from which the estimates are constructed. The index k labels the taper and n labels the trial.

	Equation	Eq.	Comment
Variance	$var\{I_{aa}^X(f)\} = \frac{2E\{I_{aa}^X(f)\}^2}{\nu(f)}$	[27]	use ν_0 for asymptotic or $\nu(f)$ if using finite size correction
Degrees of freedom	$\frac{1}{\nu(f)} = \frac{1}{\nu_0} + \frac{C_h^X \Phi(f)}{2TN_T E\{I^X(f)\}^2}$	[39]	See text for definitions of C_h^X and $\Phi(f)$
Confidence ($1 - 2p$) \times 100%	$[\nu I^X(f)/q_2, \nu I^X(f)/q_1]$	[30]	q_1 s.t $P[\chi_\nu^2 \leq q_1] = p$ q_2 s.t $P[\chi_\nu^2 \geq q_2] = p$

Table 3: Main formulae required for estimating spectral error bars. Refer to section 4 for additional information.

	Equation	Eq.	Comment
Coherency	$C^X(f) = \frac{I_{ab}^X}{\sqrt{I_{aa}^X I_{bb}^X}}$	[32]	
Distribution for coherence	$P(C) = (\nu - 2) C (1 - C ^2)^{(\nu/2-2)}$	[35]	Under null hypothesis $\gamma = 0$
Confidence for phase	$\hat{\phi}(f) \pm 2\sqrt{\frac{2}{\nu} \left(\frac{1}{ C(f) ^2} - 1 \right)}$	[36]	Approx. 95%

Table 4: Main formulae required for coherency estimation. Refer to section 5 for additional information.

	Equation	Eq.	Comment
Complex amplitude of line	$\hat{c}_1 = \frac{\sum_k J_k(f_1)H_k(0)}{\sum_k H_k(0) ^2}$	[51]	
F-test to access the significance of a line	$\frac{ \hat{c}_1 ^2 \sum_k H_k(0) ^2 (K-1)}{\sum_k J_k(f_1) - \hat{c}_1 H_k(0) ^2} \doteq F_{2,2(K-1)}$	[53]	Null $c_1 = 0$
Residual spectrum	$\hat{S}(f) = \frac{1}{K} \sum_k J_k(f) - \hat{c}_1 H_k(f - f_1) ^2$	[52]	

Table 5: Main formulae required for the detection and removal of a line from the spectrum. Refer to section 8 for additional information.

A Derivation of Finite Size Correction

The following is an outline derivation of the finite size corrections described in section 6. Firstly the characteristic functionals (Bartlett, 1966) for the processes \overline{N} and the inhomogeneous Poisson process rate $\lambda(t)$ are related.

$$C_{\overline{N}}(\theta(t)) = E\{\exp(i \int_0^T \theta(t) d\overline{N})\} = E_{\lambda}\{\exp(\int_0^T \lambda(t) b(\theta(t)) dt)\} \quad (57)$$

$$b(\theta(t)) = \exp[i\theta(t) - \frac{i}{T} \int_0^T \theta(t') dt'] \quad (58)$$

Under the Gaussian process assumption for $\lambda(t)$ this integral may be done.

$$C_{\overline{N}} = \exp[\frac{1}{2} \int_0^T \int_0^T b(t) \Lambda(t, t') b(t') dt dt' + \overline{\lambda} \int_0^T b(t) dt] \quad (59)$$

$$\Lambda(t, t') = E_{\lambda}\{(\lambda(t) - \overline{\lambda})(\lambda(t') - \overline{\lambda})\} \quad (60)$$

Note that $\overline{\lambda}$ denotes the mean rate. Taking the log of the characteristic functionals now yields the following relation between the resultant cumulant functionals.

$$K_{\overline{N}} = \ln E\{\exp(i \int_0^T \theta(t) d\overline{N})\} = \frac{1}{2} \int_0^T \int_0^T b(t) \Lambda(t, t') b(t') dt dt' + \overline{\lambda} \int_0^T b(t) dt \quad (61)$$

Next $\theta(t)$ is chosen appropriately and substituted into $K_{\overline{N}}$. The form for $\theta(t)$ which is required to obtain the covariance of multitaper estimators is;

$$i\theta(t) = \theta_1 h_k(t) e^{-2\pi i f_1 t} + \theta_2 h_k(t) e^{2\pi i f_1 t} + \theta_3 h_{k'}(t) e^{-2\pi i f_2 t} + \theta_4 h_{k'}(t) e^{2\pi i f_2 t} \quad (62)$$

Substituting into the cumulant functional for \overline{N} yields;

$$K_{\overline{N}} = \ln E\{\exp(\theta_1 J_k^D(f_1) + \theta_2 J_k^{D*}(f_1) + \theta_3 J_{k'}^D(f_2) + \theta_4 J_{k'}^{D*}(f_2))\} \quad (63)$$

Where J_k^D is the Fourier transform of the data tapered by a function indexed by k . Application of the cumulant expansion theorem (Ma, 1985) then leads to;

$$K_{\overline{N}} = E\{\exp(\theta_1 J_k^D(f_1) + \theta_2 J_k^{D*}(f_1) + \theta_3 J_{k'}^D(f_2) + \theta_4 J_{k'}^{D*}(f_2)) - 1\}_C \quad (64)$$

This may then be differentiated and set to zero.

$$K_{lmno} = \left. \frac{\partial K_{\overline{N}}}{\partial \theta_1^l \partial \theta_2^m \partial \theta_3^n \partial \theta_4^o} \right|_{\theta_1=\theta_2=\theta_3=\theta_4=0} = E\{J_k^{Dl}(f_1) J_k^{Dm*}(f_1) J_{k'}^{Dn}(f_2) J_{k'}^{Do*}(f_2)\}_C \quad (65)$$

Moments of the estimators may be expressed in terms of these cumulant derivatives. The expressions are simplified by the fact that all cumulant derivatives which have indices summing to an odd number are zero because \overline{N} is a zero mean process.

$$E\{I_k^D(f)\} = K_{1100} \quad (66)$$

$$\text{var}\{I^{MT}(f)\} = \frac{1}{K^2} \sum_{k=0}^{K-1} \sum_{k'=0}^{K-1} \text{cov}\{I_k^D(f), I_{k'}^D(f)\} \quad (67)$$

$$\text{cov}\{I_k^D(f), I_{k'}^D(f)\} = K_{1010} K_{0101} + K_{1111} + K_{1001} K_{0110} \quad (68)$$

The problem has now been reduced to that of calculating these derivatives within the model. This is done by substituting the expression for $\theta(t)$ into the RHS of equation 61. Considerable algebra then leads to the following exact result.

$$K_{lmno} = K_{lmno}^A + K_{lmno}^B \quad (69)$$

Where,

$$K_{lmno}^A = \frac{1}{2} \sum_{l_i, m_i, n_i, o_i} \frac{l!m!n!o!}{\prod l_i! \prod m_i! \prod n_i! \prod o_i!} \left[\frac{-H_1(f_1)}{T} \right]^{l_2+l_4} \left[\frac{-H_1(f_1)^*}{T} \right]^{m_2+m_4} \dots \left[\frac{-H_1(f_2)}{T} \right]^{n_2+n_4} \left[\frac{-H_1(f_2)^*}{T} \right]^{o_2+o_4} I_{l_3, m_3, n_3, o_3}^{l_1, m_1, n_1, o_1} \quad (70)$$

Where $\sum_i l_i = l$ and cases where $l_1 + l_2 = l$ or $l_3 + l_4 = l$ are excluded (and also for n, m, o).

$$I_{l_3, m_3, n_3, o_3}^{l_1, m_1, n_1, o_1} = \int_{-\infty}^{\infty} S_{\lambda}(f) H_{l_1+m_1+n_1+o_1}[f_1(l_1 - m_1) + f_2(n_1 - o_1) - f] \dots H_{l_3+m_3+n_3+o_3}^*[f_1(l_3 - m_3) + f_2(n_3 - o_3) - f] df \quad (71)$$

Where $S_\lambda(f)$ is the spectrum of the Gaussian process and H_l is;

$$H_l(f) = \int_{-\infty}^{\infty} h^l(t) \exp(-2\pi i f t) dt \quad (72)$$

$$H_0(f) = T \exp(-i\pi f T) \text{sinc}(\pi f T) \quad (73)$$

$$K_{lmno}^B = \bar{\lambda} \sum_{p=0}^l \sum_{q=0}^m \sum_{r=0}^n \sum_{s=0}^o [\begin{smallmatrix} l \\ p \end{smallmatrix}] [\begin{smallmatrix} m \\ q \end{smallmatrix}] [\begin{smallmatrix} n \\ r \end{smallmatrix}] [\begin{smallmatrix} o \\ s \end{smallmatrix}] H_{p+q+r+s} [f_1(p-q) + f_2(r-s)] \dots$$

$$\left[\frac{-H_1(f_1)}{T} \right]^{(l-p)} \left[\frac{-H_1(f_1)^*}{T} \right]^{(m-q)} \left[\frac{-H_1(f_2)}{T} \right]^{(n-r)} \left[\frac{-H_1(f_2)^*}{T} \right]^{(o-s)} \quad (74)$$

The preceding result is somewhat cumbersome but readily evaluated computationally for a given spectrum. The expression simplifies greatly when only frequencies above the smoothing width are considered and many of the terms may be neglected. Restricting attention to the second order properties there are only a few remaining dominant terms. Terms from K_{1001} lead to the previously discussed asymptotic results but there are corrections which arise from the term K_{1111} . Assuming that the population spectrum varies slowly over the width of the tapers leads to the result given by equations 38 - 43. The validity of this assumption has been tested computationally and was found to be very accurate even for spectra with sharp peaks.

Acknowledgment

The authors thank C. Buneo, and R. Sachdev for providing example datasets, Clive Lauder for help with the calculation of global error bars and D. R. Brillinger and D. J. Thomson for comments which substantially improved the manuscript. M. Jarvis is grateful to R. A. Andersen both for his continued support of theoretical work in his lab and also for his careful reading of the manuscript. M. Jarvis acknowledges the generous support of the Sloan foundation for theoretical neuroscience.

References

Abeles, M., Deribaupierre, F., and Deribaupierre, Y. (1983). Detection of single unit responses which are loosely time-locked to a stimulus. *IEEE Transactions on Systems Man and Cybernetics*, 13:683–691.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley.

- Bartlett, M. S. (1966). *An Introduction to Stochastic Processes*. Cambridge University Press.
- Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. 6th Berkley Symposium on Mathematical Statistics and Probability*, 1:483–513.
- Brillinger, D. R. (1974). *Time Series*. Holt, Rinehart and Winston Inc.
- Brillinger, D. R. (1978). *Develpomentns in Statistics vol1*, chapter Comparative Aspects of the Study of Ordinary Time Series and of Point Processes, pages 33–129. Academic Press. Inc.
- Brody, C. D. (1998). Slow covariations in neuronal resting potentials can lead to artefactually fast cross-correlations in their spike trains. *Journal of Neurophysiology*, 80:3345.
- Cox, D. R. and Lewis, P. (1966). *The Statistical Analysis of Series of Events*. Chapman and Hall.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- Gerstein, G. L., Perkel, D. H., and Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously recorded population of neurons - detection. *Journal of Neuroscience*, 4:881–889.
- Gray, C. M., Konig, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual-cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337.
- Hannan, E. J. (1970). *Multiple Time Series*. Wiley.
- Ma, S. K. (1985). *Statistical Mechanics*. River Edge : World Scientific Publishing Company.
- Mitra, P. P. and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal*, 76:691–708.
- Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.
- Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (submitted 2000). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *J. Neuroscience*.

- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes*. Cambridge University Press.
- Rosenberg, J. R., Amjad, A. M., Breeze, P., Brillinger, D. R., and Halliday, D. M. (1989). The fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. Molec. Biol.*, 53:1–31.
- Sachdev, R., Jenkinson, E., Melzer, P., and Ebner, F. F. (1999). Dual electrode recordings from the awake rat barrel cortex. *Somatosensory and Motor Research*, 16:163–192.
- Snyder, L. H., Batista, A. P., and Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, 386:167–169.
- Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22:1328–1345.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of IEEE*, 70:1055–96.
- Thomson, D. J. and Chave, A. D. (1991). *Advances in Spectrum Analysis and Array Processing*, chapter Jackknifed Error Estimates for Spectra, Coherences, and Transfer Functions, pages 58–113. Prentice Hall.